COMP S450 Applied Computing Project

DREAMS: A ONE-STOP DOCUMENT RECOGNITION, EXTRACTION AND MANIPULATION SOLUTION FOR LOGISTICS INDUSTRY

Group members:

1155180	Eraka Yip	(BCOMPHIT1)
1265005	Timothy Chan	(BSCHCN)
1284810	Andy Lau	(BCOMPHIT1)

Project Supervisor: Dr. Voyce Li

Course Coordinator: Dr. Jeff Au Yeung

Table of Content

1. Introduction

1.1. Background

In the logistics industry, the sea transport sector is very important in supporting Hong Kong as the world's sixth largest trading entity. Exports and imports are the major part of the sea transport sector, so Hong Kong has been committed to promoting smart ports and optimizing import and export services.

As part of the port community, our company also aims to achieve this. Therefore, our company provides an import and export platform that can be used by different stakeholders and achieves paperless import and export services.

Although the platform has been officially used, there are still some unresolved problems. In order to understand the problems more clearly, we first introduce the stakeholders of the platform.

Stakeholders	Description
Shipping Line	• Provide vessels for sea transport services, work with terminals to loading or discharge containers. • Provide data for terminal operation.
Terminal	• Control gate in and out of a port. • Provide container storage services for pickup and return containers. • Provide data for the platform to show information for users (Shipper and Trucking Company).
Platform Provider, our company	• Provide a platform that users (shipper and trucking company) can complete export and import services without using paper documents.
Shipper	• Request Booking from Shipping Line, then use the platform to retrieve booking data and assign jobs to trucking companies. ** Since booking data from the shipping line does not include shipper information, shippers need to retrieve booking from the platform themselves.
Trucking Company	• Receive job from shipper, then assign drivers to pick up or return containers. • Drivers can use a mobile app which is provided by platform to gate in and out terminals instead of using paper documents.

Next, the main workflow of the import and export platform will be shown below:

1. Retrieve Booking (Shipper) ⬚ Problem found in this step, detail please found below figure.
2. Forward Booking to Trucking Company (Shipper)
3. Assign Driver to pick up container from terminal (Trucking Company)
4. After container picked and cargo packed into container, submit declaration for return containers to terminal (Shipper, Trucking Company)
5. After the container is returned, get the transshipment receipt, and wait for container loading on the vessel. (Shipper)
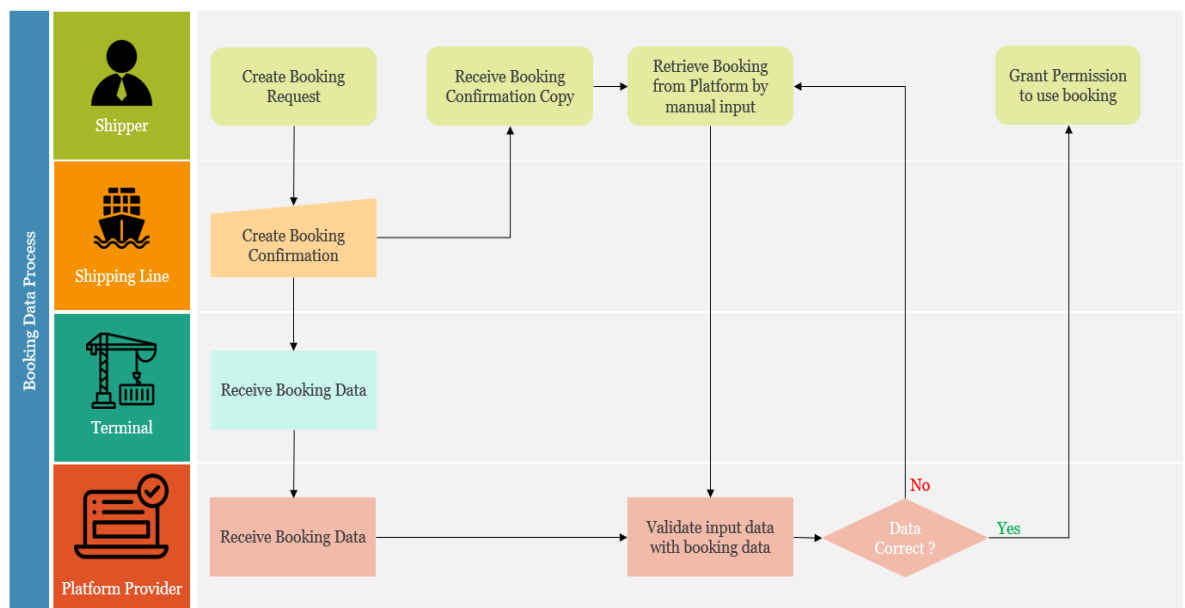

Figure 1: booking retrieval data process flow

1.2. Project Aim

Problem in Current Workflow
According to the current workflow, after booking is retrieved by a user successfully, other users cannot use the same data. In addition, in the past, manual inspection of documents was required for entry and exit at the terminal gate. This is no longer required. Instead, the platform provides consolidated data to the terminal for review, and users only need to use the platform's mobile app (as convenient as a student scanning a student ID card when returning to school), saving the time and human error of manual inspection. In this way, data security and workflow digitizing can be taken into consideration at the same time.

Disadvantage in Current Workflow
Although the current workflow seems good, we can find out the problem in the retrieve booking process, since the shipping line doesn't provide customer information to us, so we can't digitize the workflow and validate the access rights of the retrieve user.

Now we assume only the booking owner can get the booking confirmation copy and know the booking number, so we require the user to manual input booking number to retrieve booking.

Let's list out disadvantage of this solution:

1. Low data security - Booking Number is sequential for each shipping line, so the user can guess other valid booking numbers easily. Although they can't steal the booking actually, but the booking data contains sensitive data on it, disclosing those data will lower data security.

2. Non-digital workflow - Manual input increases user workload, reduces efficiency and increases the risk of manual input errors, in addition to lowering user experience.

Based on the above disadvantage, we proposed a document processing solution that leverages Optical Character Recognition (OCR) and machine learning methodologies to enhance data security and reduce manual, repetitive input tasks among collaborating parties.

1.3. Project Objectives

1. Collect data and build dataset
   - Five document formats are considered in our projects. We will prepare about 30 documents per format.

2. Setup AI training environment
   - Prepare required techniques listed in chapter 3, including basic machine learning environment, additional pre-trained models and the libraries that the models required.

3. Train, test and evaluate models
   - In our preliminary plan, there are two types of models in our system for document classification and data extraction.

4. Select the best models
   - Since there are more than one pre-trained models to achieve our goals, the model with best performance should be selected after complete training in each type.

5. Package and evaluate the system
   - Besides linking up the two selected models, we should also create interfaces for submitting documents and existing logistic procedures. Finally, the whole system will be evaluated.

1.4. Value Propositions

If this project is successful, it will improve data security and protect user data from leaks. Preparability will also be improved by being able to cross-check whether the data in the

database and the uploaded file are the same. Reducing manual input will also increase work efficiency.

In the long term, this project can contribute to the complete digitization of the entire import and export workflow. And by bringing more data to the company, we can look forward to becoming a data-driven company in the future, making different analyzes for customers and increasing revenue.

2. Background or Literature Review

2.1. Review of Existing or Related Solutions for the Problem

Our Company has proposed some solutions for this problem. First, we request the shipping line to send a unique ID to both shipper and terminal, then the terminal passes the unique ID to our database and the shipper can use this unique ID to retrieve booking from our platform. This method ensures data security and keeps workflows simple, but this requires the shipping line to change their system, since shipping line is a global company which is very difficult to change their system for another local company, and Hong Kong has so many shipping lines, this approach is not realistic.

Next, we request the shipping line to provide a field for shipper input their company ID when request booking and send to the terminal. This method only needs the shipping line to add an input which is less development effort, but the problem is we are difficult to control if the shipper inputs the company ID incorrectly.

Last, we request the shipper to input more information to retrieve the booking. This method does not need any shipping line enhancement, and also keeps data security since multiple booking information is very difficult to guess. But it makes the workflow complicated, so it is very difficult to retrieve a booking, so this method does not work.

2.2. Related Supporting Technologies

Given the current situation, we searched for various supporting technologies that can give us some insights on the project direction.

1. Document processing SaaS / software
   a) SaaS solution from public cloud providers.
   Nowadays, many companies consider migrating their on-premises data centers to the cloud to enjoy the benefits of cloud services. Companies can pay based on usage, enjoy high availability services agreed in SLA, and can scale with high elasticity. This can help companies reduce their maintenance cost and can respond to sudden demand change quickly.

   Likewise, the big three cloud providers – Amazon web services, Microsoft Azure and Google Cloud Platform (GCP) are also offering their intelligent document processing solution in software-as-a-Service (SaaS). Users upload the documents to the cloud. The SaaS solution will analyze the document layout and extract the key-value pairs based on their pre-built AI models. We can train our custom model to achieve better performance and fit the specified use case.

   b) AI-powered document processing software
   Meanwhile, there are several AI-powered document processing software available in the market. This kind of software makes use of their custom AI models to analyze the

document and extract the key-value pairs. Studies showed some can outperform current available AI models.

2. Robotic Process Automation (RPA)
   RPA mimics human actions such as keyboard and mouse operations. With OCR tool and specified UI actions, it can analyze the key-value pair from the document with certain layouts. It is scalable and can run without human intervention (unattended RPA).

   Common RPA software include Uipath, Microsoft Power Automate and Automation anywhere.

However, there are several considerations when using these technologies as an enterprise solution.

a) Data governance policy
   We have to update our client's secured data to the public cloud for document processing, which may not be allowed in IT policy.

b) Black Box nature of market solutions
   We have no idea of the details of the AI models, including the algorithm and exact structure of their models. We also lose chances to learn new technology and coding skills.

c) Lack of control
   We expect to have better control on the model. However, in a SaaS solution, the provider is responsible for providing the updates and patches. We may need to verify the performance of the model afterwards.

We observed that these technologies are highly equipped with AI and it seems to be a solution to the current problem.

In recent years, studies on machine learning have been greatly shifted to deep learning, which is a subset of machine learning using artificial neural networks such as CNN and RNN for tasks such as object detection and layout analysis. They achieve outstanding results in general.

However, there are some constraints on building these neural models - To train the model, we need to input a large (labeled) dataset with powerful GPUs and long training time.

Fortunately, many big techs have developed their open-source, transformer-based model. where we can use a smaller dataset for tuning the big, pre-trained model instead of building from scratch.

For example, YOLO for image detection. We found that there were three FYPs last year using YOLO for their object detection model.

Also, DocBERT for document classification and LayoutLM and Donut for document layout analysis. These are potential models that can be employed in our project.

2.3. Conclusion

We decided to develop custom AI models for the project based on pre-trained transformer-based models.

COMP S450 Applied Computing Project

3. Preliminary Methodology

3.1. Overview

In the new system, shipper send the document to us by email. Then, we will classify the document by the owner and extract required data from the documents. Finally, the extracted data will become the input of our existing logistics procedure.
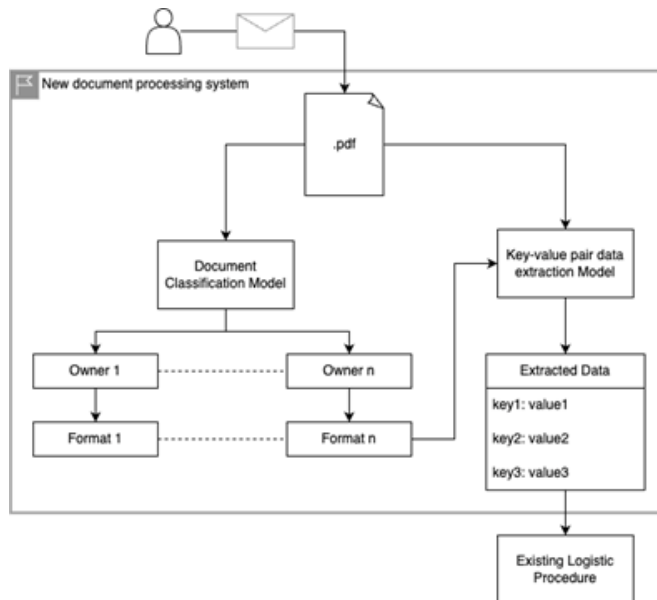

Figure 2: The system architecture of our document processing system

3.2. Requirements, Supporting Technologies, and Technical Gap
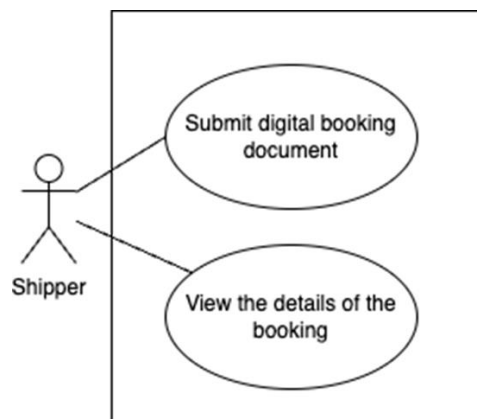
3.2.1. Requirements

Figure 3: A use case diagram of the system

User Types	Functions
Shipper	Submit digital booking document View the details of booking

The stuff of shipper companies submit the digital booking documents to our new system by email. After the extracted data is stored in the database, the staff can view the details of booking and do further action.

Functional requirements

1. Verify the documents
   Documents with invalid format should be rejected to increase the accuracy and improve the security.

2. Classify the document into different shipping line companies
   It is an Important step to extract document data from a certain company.

3. Extract data from the document
   Extract essential data from the documents to replace manual inputs.

Non-functional requirements

1. Accuracy
   This is important to extract correct data from the documents in order to reduce human workload effectively. Therefore, good quality of datasets is required to classify the document and hence extract essential data correctly.

2. Scalability
   Our systems support only five document formats. However, there are 60 shipping line companies in Hong Kong. To make our system practical, scalability is required for further implementation. In the end, this system should completely accept all booking documents in Hong Kong. Moreover, the formats of the documents may be updated after a period of time. The supported formats in our system keep rising.

3. Performance
   All the shippers in Hong Kong use this system to arrange logistics tasks. Enormous bookings are retrieved every day. Higher performance increases the capacity and improves the user experience.

3.2.2. Supporting technologies

1. Python
   Most popular programming language used in AI.

2. Machine learning
   The machine "learns" the patterns from provided datasets.

3. Documents classification models
   The popular pre-trained models to classify documents by visual characteristics (YOLO) or content (DocBERT).

4. Key-value extraction models
   The popular pre-trained models extract data by trained region of interest.

3.2.3. Technical gap

1. Lack of experience in training AI model
   All of our team members haven't prepared training datasets, especially high quality training data to perform better performance.

2. Implementing various pre-trained models
   All the pre-trained models to be used in our project are new to us. In the following time, we should understand their requirements and fine-tuning implementations.

3. Choosing the suitable development tools and environments
   Until now, we are still suffering in choosing development environments from cloud services or local devices.

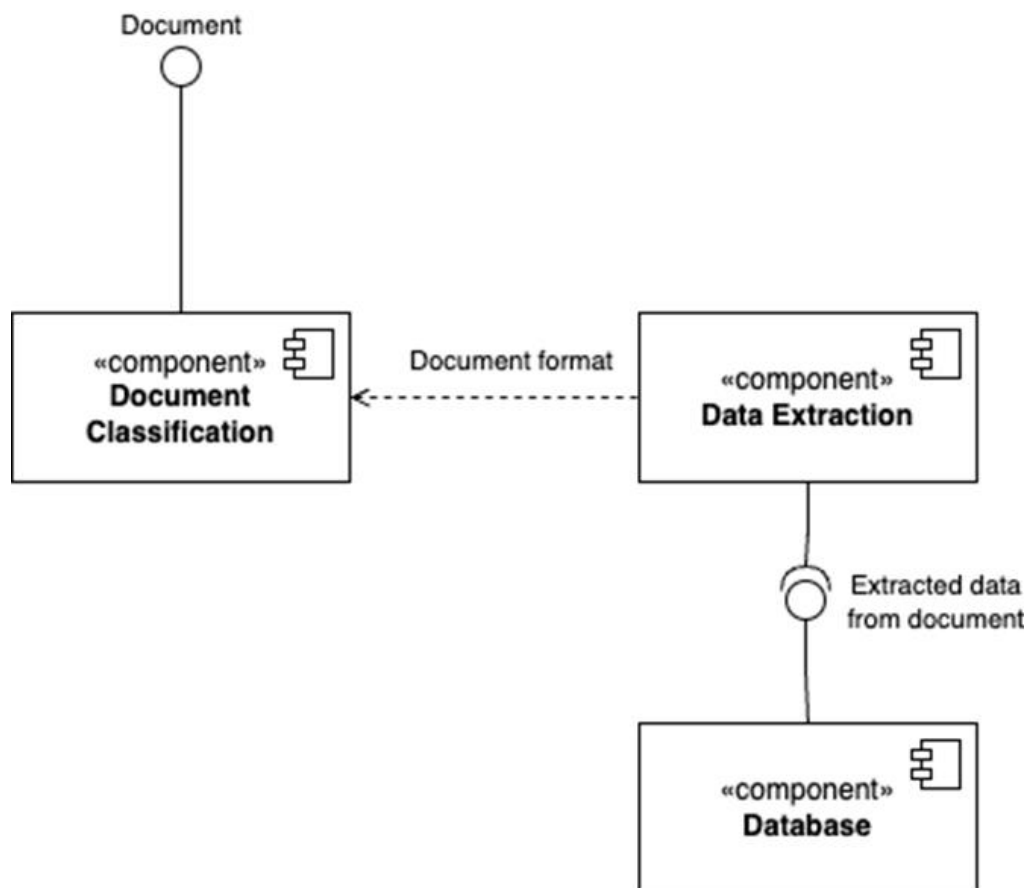## 3.3. Architecture or High-Level System Design

Figure 4: Component diagram

The Document Classification component provides an interface to receive digital documents. Data extraction component depends on the result of classification. The booking number

extracted from the document will be stored under the user who submit the document through database interface.
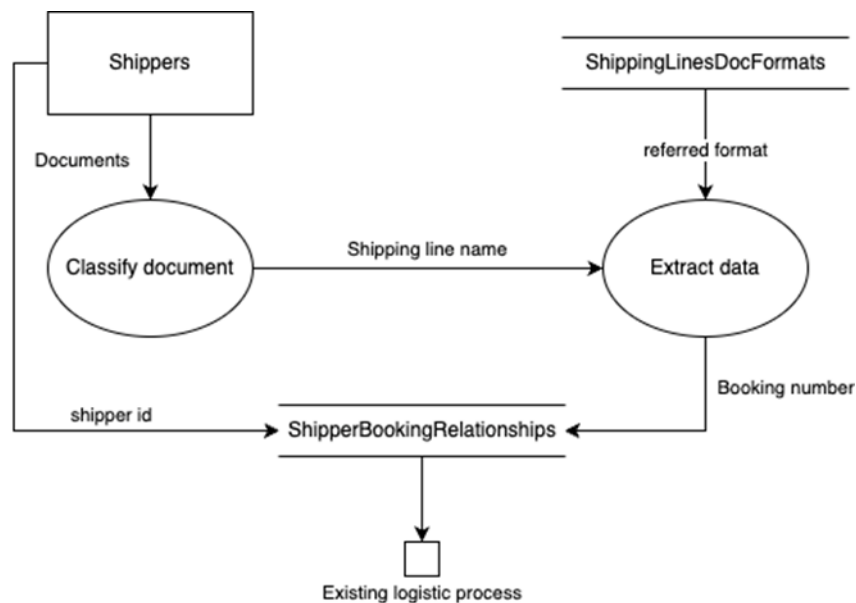

Figure 5: Data flow diagram

The documents provided by shippers will be classified to output shipping line name to data extraction process. The weight file (format) will be obtained from ShippLinesDocFormats data store by shipping line name to extract data. The relationship between bookings and shippers will be created by the booking number extracted and the shipper id. The relationship will be used in the existing logistics process out of our system.

The major components

Component 1 - Documents classification

We planned to classify the documents by their characteristics. For example, the company logo on the document can identify which shipping line companies the document belongs to. They can be detected by YOLO. Also, the content of the document can be used to analyze whether the document is booking or not by transformer-based models such as DocBERT.

Component 2 - Data extraction

After we find out the owner of the documents, essential data can be extracted from the received documents. There are two potential pre-trained large models we are going to use, which are LayoutLM and Donut. This is important to choose a model to provide better accuracy. Each shipping line company has their own trained model.

References

Adhikari, A., Ram, A., Tang, R., & Lin, J. (2019). Docbert: Bert for document classification. arXiv preprint arXiv:1904.08398.

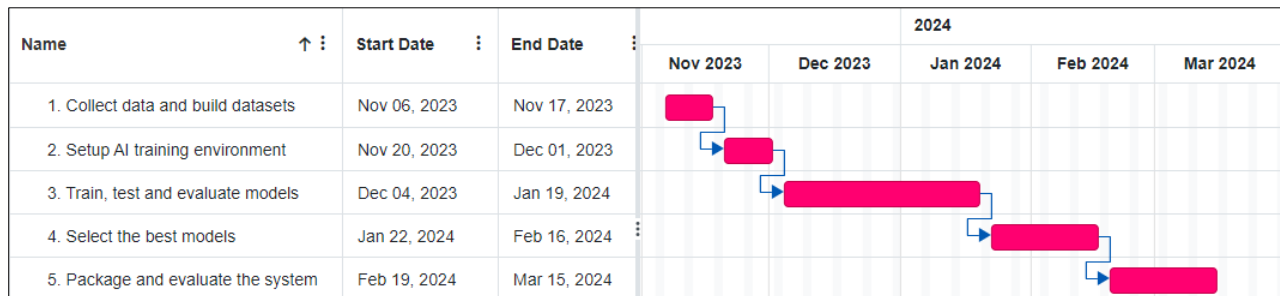Bhatt, A. (2022) Document Automation Using Artificial Intelligence.

Kim, G., Hong, T., Yim, M., Park, J., Yim, J., Hwang, W., ... & Park, S. (2021). Donut: Document understanding transformer without ocr. arXiv preprint arXiv:2111.15664, 7, 15.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020, August). Layoutlm: Pre-training of text and layout for document image understanding. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1192-1200).

Appendix A. Project Plan

Gantt Chart

Name ↑	Start Date	End Date	2024				
			Nov 2023	Dec 2023	Jan 2024	Feb 2024	Mar 2024
1. Collect data and build datasets	Nov 06, 2023	Nov 17, 2023					
2. Setup AI training environment	Nov 20, 2023	Dec 01, 2023					
3. Train, test and evaluate models	Dec 04, 2023	Jan 19, 2024					
4. Select the best models	Jan 22, 2024	Feb 16, 2024					
5. Package and evaluate the system	Feb 19, 2024	Mar 15, 2024					

Please refer to section 1.3 Projective objective for the details of task.

Appendix B. Team Members' Roles and Responsibility

Roles

Roles	Member(s)	Remarks
Team coordinator	Eraka Yip	Manages the project in general and keeps records, reports, and other documents in order, and prepare the submission of reports
Team members (Roles: Designer, Programmer, Tester and System Analyst)	Eraka Yip, Timothy Chan, Andy Lau	All team members will participate in designing, coding, testing and deploying the system.

Responsibilities and Task Assignment

Tasks	Responsible Member(s)	Target Date
Collect data and build dataset	Eraka Yip, Timothy Chan, Andy Lau	Nov 17, 2023
Setup AI training environment	Eraka Yip, Timothy Chan, Andy Lau	Dec 01, 2023
Train, test and evaluate models	Eraka Yip, Timothy Chan, Andy Lau	Jan 19, 2024
Select the best models	Eraka Yip, Timothy Chan, Andy Lau	Feb 16, 2024
Package and evaluate the system	Eraka Yip, Timothy Chan, Andy Lau	Mar 15, 2024

Remarks: Tasks such as report writing and presentation demonstration are not included in the Gantt chart since they are not our project objectives. However, they are work-intensive and crucial for the success of our project.